

多変量解析を用いた力学系の埋め込み次元解析

The embedding dimension analysis of dynamical
systems by multivariate analysis

中央大学大学院理工学研究科博士前期課程物理学専攻
谷本智也

2004年3月11日

目次

第 1 章	はじめに	2
第 2 章	埋め込み定理	3
第 3 章	主成分分析	7
3.1	主成分分析とは	7
3.2	主成分分析の定式化	8
3.3	寄与率, 累積寄与率	9
3.4	多次元の場合について	9
3.5	主成分分析の例	10
第 4 章	正準相関分析	12
4.1	正準相関分析とは	12
4.2	正準相関分析の定式化	13
4.3	正準相関分析の例	15
第 5 章	数値計算	20
5.1	主成分分析の結果	20
5.2	正準相関分析の結果	26
第 6 章	まとめ	28

第1章 はじめに

ターケンスの埋め込み定理により, m 次元の力学系の1つの変数 $y(t)$ の時系列データから適切な時間隔 Δt をとり作られた k 次元の時間遅れベクトル

$$\vec{v}(t) = (y(t), y(t + \Delta t), \dots, y(t + (k - 1)\Delta t))$$

が $k > 2m$ であるならば時間遅れベクトルから再構成されたアトラクターは k 次元の空間に埋め込まれるというのはよく知られている。

しかし適切な時間間隔を見つける定性的な方法はなく, を見つけることは困難である。またこの埋め込み定理では, たかだか k 次元の空間に埋め込まれることだけが保証され, その力学系の次元を特定することは原理上できない。

このように時間遅れベクトルからアトラクターを再構成することは, 一般的に使われているが, より精度のいい観測をするには更なる手法を開発をする必要がある。

この研究の目的は, 多変量解析の手法である主成分分析, 正準相関分析をもちいて1つの時系列データからアトラクターを再構成し, 元の力学系の次元を推定する手法を開発することである。主成分分析はより少ない次元に情報を縮減する手法で, 縮減された次元から力学系の次元を推定していく。また正準相関分析は2つのグループの相関を調べる手法で, これを用いて主成分分析で再構成されたアトラクターの成分が元の力学系のどの成分から構成されているかを調べる。

第2章 埋め込み定理

m次元の連続時間力学系

$$\frac{d}{dt}\vec{x}(t) = \vec{f}(\vec{x}(t))$$

に初期条件が与えられ、十分時間がたった後の、m次元状態空間内での漸近的な振る舞いを力学系のアトラクターという。アトラクターは力学系の幾何的な特徴を表している。

力学系のアトラクターは

1. 平衡点 (fixed point)(図 2.1)
2. リミットサイクル (limit cycle)(図 2.2)
3. トーラス (torus)(図 2.3)
4. ストレンジアトラクター (strange attractor) またはカオス (chaos)(図 2.4)

に分類される。

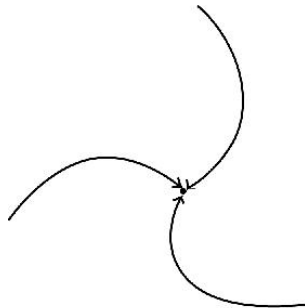


図 2.1: 平衡点

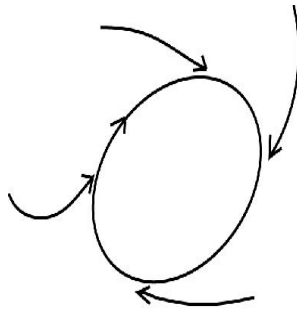


図 2.2: リミットサイクル

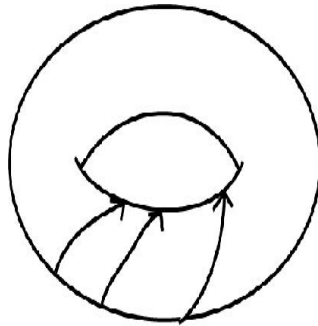


図 2.3: トーラス



図 2.4: ストレンジアトラクター

しかし与えられた力学系において、実際の実験では m 次元空間における全状態変数 \vec{x} を完全に観測することはほとんどできず、最悪の場合 1 変数の時系列データ $y(t)$ が観測できるのみである。このような場合、観測された 1 次元のデータから高次元空間における力学系のアトラクタの軌道を再構成する必要がある。このアトラクタの再構成に最も良く使われている手法は、時間遅れ座標系への変換である。

この手法では、時間遅れの間隔を Δt として、 k 次元の再構成状態空間において以下のような k 次元の時間遅れベクトル

$$\vec{v}(t) = (y(t), y(t + \Delta t), \dots, y(t + (k - 1)\Delta t))$$

を作りこの時間遅れベクトルを k 次元の空間に構成していけば k 次元のアトラクタが再構成される。

3 次元空間に時間遅れベクトルが埋め込まれアトラクタが再構成されるイメージを図 2.5 で示す。

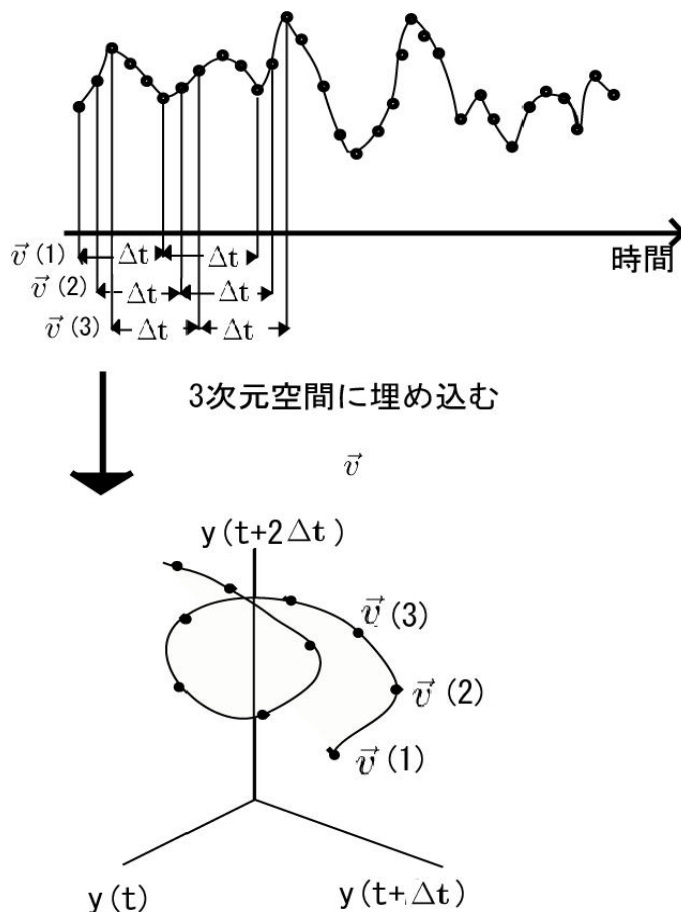


図 2.5: 埋め込みのイメージ

こうして再構成されたアトラクタの再構成状態空間の次元 k が元の力学系の次元 m に対して $k > 2m$ ならば、観測された時系列から再構成次元への変換が埋め込みであることが以下の定理で示されている。

定理 1 (F.Takens の埋め込み定理) m 次元のコンパクトな多様体 M と C^2 級写像 $f: M \mapsto M, g: M \mapsto R^1$ が与えられたとき、 $k > 2m$ であれば、以下の写像 $v: M \mapsto R^k$ は埋め込みである。

$$\vec{v}(x) = (g(x), g(f(x)), g(f^2(x)), \dots, g(f^{k-1}(x)))$$

ここで埋め込みとは、写像が 1 対 1 かつはめ込みであることであり、はめ込みとは x における f の微分が 1 対 1 である事である。

この定理では適当な Δt をとれば、トポロジーが保たれるが次元に関してはたかだか $k > 2m$ が保証されているだけである。

第3章 主成分分析

3.1 主成分分析とは

主成分分析は例えば

- n 人の患者についての発熱, 腹痛, … などの p 種類の症状の強さ
- 企業 n 社についての利益率, 自己資本比率 … などの p 種類の財務指標の値

等の高い次元のデータを, 情報をできるだけ失うことなく 1 個または少数の次元へ要約する多変量解析の 1 つの方法である。各変数を線形結合し, 分散が最大になるように係数を決める。このように決めた線形結合を新しい変数とすればより多くの情報を 1 つの変数に縮約できる。図 3.1 のように変数間 (x, y) に大きな相関があれば新しい変数 (z) でより大きな情報を表すことができる。

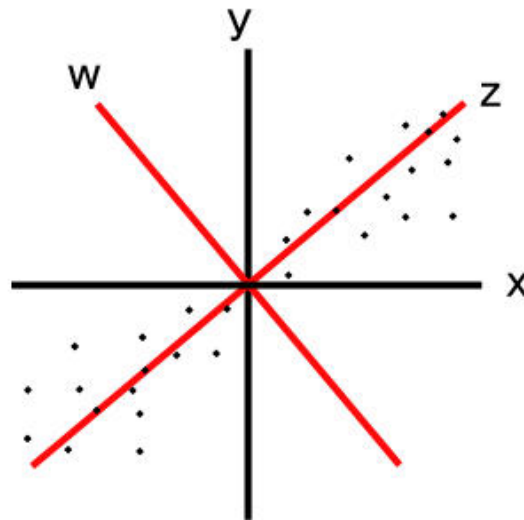


図 3.1: 主成分分析のイメージ

3.2 主成分分析の定式化

ここでは2次元について説明するがより高次元の場合でも本質的に変わらない。
2つの変数 X, Y があり2つの変数の線形結合

$$Z = \alpha_1 X + \alpha_2 Y = \vec{\alpha}' \vec{w}$$

の分散

$$\sigma_Z = \alpha_1^2 \sigma_{11} + 2\alpha_1 \alpha_2 \sigma_{12} + \alpha_2^2 \sigma_{22} = \vec{\alpha}' \Sigma \vec{\alpha}$$

を最大にするように係数 $\vec{\alpha}$ を決める。ここで σ_h は h の分散, σ_{ij} は i と j の共分散, Σ は分散共分散行列である。 $\vec{\alpha}$ はこのままでは解が不定なので制約条件として

$$\alpha_1^2 + \alpha_2^2 = \vec{\alpha}' \vec{\alpha} = 1$$

を与える。この時 $\vec{\alpha}$ は Lagrange の未定乗数法により

$$\phi(\vec{\alpha}) = \alpha_1^2 \sigma_{11} + 2\alpha_1 \alpha_2 \sigma_{12} + \alpha_2^2 \sigma_{22} - \lambda(\alpha_1^2 + \alpha_2^2 - 1)$$

を α_1, α_2 で微分して0でくと

$$\left. \begin{aligned} (\sigma_{11} - \lambda)\alpha_1 + \sigma_{12}\alpha_2 &= 0 \\ \sigma_{21}\alpha_1 + (\sigma_{22} - \lambda)\alpha_2 &= 0 \end{aligned} \right\} \\ \Rightarrow (\Sigma - \lambda I_2) \vec{\alpha} = 0$$

となる。この方程式の $\vec{\alpha}$ が自明でない解を持つための必要, 十分条件は

$$|\Sigma - \lambda I_2| = 0$$

である。つまり, λ は Σ の固有値であり, $\vec{\alpha}$ は対応する固有値ベクトルである。この固有値 $\lambda_1, \lambda_2 (\lambda_1 > \lambda_2)$ に対応する固有ベクトル $\vec{\beta}_1, \vec{\beta}_2$ とすれば

$$\Sigma \vec{\beta}_i = \lambda \vec{\beta}_i$$

なので, 両辺に $\vec{\gamma}'$ を左から掛けると制約条件 $\vec{\alpha}' \vec{\alpha} = 1$ より

$$\vec{\beta}_i' \Sigma \vec{\beta}_i = \vec{\beta}_i' \lambda \vec{\beta}_i = \lambda_i$$

となり固有値 λ_i は分散となる。

したがって分散を最大にする線形結合は, λ_1 に対応する固有ベクトル $\vec{\beta}_1$ の要素を係数とすればよい。

こうしてできた \vec{w} の線形結合

$$Z_1 = \vec{\beta}_1' \vec{w} = \beta_1 X + \beta_2 Y$$

を第1主成分と呼ぶ。第1主成分で説明しきれない部分を説明するため新しい線形結合である第2主成分

$$Z_2 = \vec{\gamma}' \vec{w}$$

を考える。第 1 主成分で説明できない部分を説明するので第 2 主成分は第 1 主成分と無相関つまり直交する必要がある。これを求めるには直交条件

$$\vec{\gamma}'\vec{\beta}_1 = 0$$

のもとでラグランジュの未定乗数法を解く。

$$\varphi = \vec{\gamma}'\Sigma\vec{\gamma} - \lambda(\vec{\gamma}'\vec{\gamma} - 1) - \nu(\vec{\gamma}'\vec{\beta}_1)$$

これを $\vec{\gamma}$ の各要素 γ_1, γ_2 で偏微分して 0 と置くと

$$\Sigma\vec{\gamma} - \lambda\vec{\gamma} - \frac{\nu}{2}\vec{\beta}_1 = 0$$

この両辺に、左から $\vec{\gamma}'$ をかけ $\vec{\gamma}'\vec{\beta}_1 = 0$ を考慮するとこの式は、第 1 主成分を求めたのと同じ固有値問題の式に帰着する。 λ_1 は第 1 主成分に用いているので、第 2 主成分では固有値、固有ベクトルに $\lambda_2, \vec{\beta}_2$ を用いればよい。

3.3 寄与率，累積寄与率

こうして求められた主成分の分散と、元の変数の分散の間には、

$$\lambda_1 + \lambda_2 = \sigma_{11} + \sigma_{22}$$

のような関係があり、

$$\frac{\lambda_i}{\lambda_1 + \lambda_2}, (i = 1, 2)$$

を第 i 主成分の寄与率という。各寄与率を足し合わせたものを累積寄与率という。

寄与率はその主成分がどの程度、情報を表しているかの尺度である。

3.4 多次元の場合について

3 次元以上の多次元 s の場合も同様に線形結合

$$Z = \alpha_1 X + \alpha_2 Y + \cdots + \alpha_s V$$

の分散から固有値問題に帰着させ主成分の係数を決めていく。そして累積寄与率が十分大きくなるまで第 1, 第 2, 第 3, \cdots と主成分を決めていく。

3.5 主成分分析の例

主成分分析の例として中学 1 年生の身体測定データのデータを分析する。30 人の中学 1 年生の身長，体重，胸囲，座高のデータを以下で示す。

生徒番号	身長 (cm)	体重 (kg)	胸囲 (cm)	座高 (cm)
1	148	41	72	78
2	160	49	77	86
3	159	45	80	86
4	153	43	76	83
5	151	42	77	83
6	140	29	64	74
7	158	49	78	83
8	137	31	66	73
9	149	47	82	79
10	160	47	74	87
11	151	42	73	82
12	157	39	68	82
13	157	48	80	88
14	144	36	68	76
15	139	32	68	73
16	139	34	71	76
17	149	26	67	79
18	142	31	66	76
19	150	43	77	79
20	139	31	68	74
21	161	47	78	84
22	140	33	67	77
23	152	35	73	79
24	145	35	70	77
25	156	44	78	85
26	147	38	73	78
27	147	30	65	75
28	151	36	74	80
29	141	30	67	76
30	148	38	79	78
平均	149	38.700	72.233	79.367
分散	53.524	41.732	26.533	18.241

これより分散共分散行列は

$$\Sigma = \begin{pmatrix} 53.524 & 40.794 & 27.589 & 28.761 \\ 40.794 & 41.732 & 29.832 & 24.355 \\ 27.589 & 29.832 & 26.531 & 17.223 \\ 28.761 & 24.355 & 17.233 & 18.241 \end{pmatrix}$$

この分散共分散行列から，固有値問題を解き主成分を求める。以下が求めた主成分の係数と寄与率である。

	第 1 主成分	第 2 主成分	第 3 主成分	第 4 主成分
係数 (身長)	0.4970	-0.5432	-0.4495	-0.5058
係数 (体重)	0.5146	0.2102	-0.4625	0.6907
係数 (胸囲)	0.4809	0.7246	0.1753	-0.4614
係数 (座高)	0.5069	-0.3683	0.07939	0.2325
寄与率	0.8853	0.0783	0.0199	0.0165
累積寄与率	0.8853	0.9636	0.9835	1.0000

第 1 主成分の寄与率が約 0.9 なので，第 1 主成分だけでこの中学生の体格についての情報はほとんど表せている。

第 1 主成分では，係数はいずれも正で 0.5 前後の値になっており，合計値の定数倍という形にほぼ近い。これはどの変数値が大きくなってもこの主成分の値は大きくなるから，第 1 主成分は全体的な大きさを表していると解釈される。

第 1 主成分まででほとんどの情報を表しているがここではより詳しく調べるため第 2 主成分についても調べる。

第 2 主成分では係数は体重と胸囲で正，身長と座高で負だから，背が低く体重の重い生徒では大きい値に，背が高く体重が軽い生徒では小さい値になる。したがって第 2 主成分は太っているかやせているかを表す主成分と解釈される。

第4章 正準相関分析

4.1 正準相関分析とは

例えば

- 主要な生産物の価格（第1群）と生産量（第2群）
- ある疾患の患者のいろいろな症状の強さ（第1群）と物理化学的な検査の成績（第2群）

のような2組の変数のグループがあり，これら変数のグループ間の相関関係を分析したいとき用いるのが正準相関分析である。

正準相関分析では， q 個と p 個の変数を持つ2つの変数のグループ

$$\vec{X}' = (x_1, x_2, \dots, x_q)$$

$$\vec{Y}' = (y_1, y_2, \dots, y_p)$$

の線形結合

$$f = a_1x_1 + a_2x_2 + \dots + a_qx_q = \vec{a}'\vec{X}$$

$$g = b_1y_1 + b_2y_2 + \dots + b_py_p = \vec{b}'\vec{Y}$$

の相関係数が最大にするような係数ベクトル \vec{a}, \vec{b} を求める。こうしてできた線形結合を正準変量と言う。図 4.1 のように \vec{X} と \vec{Y} のつくる空間 $s(X), s(Y)$ での f と g の角度が最小のとき，相関係数が最大になる。

求めた正準変量と元の変数との共分散を正準負荷量といい，各変数とその正準変量によってどのくらい代表されるかを表している。

取り出された正準変量がどれだけの情報を有しているのかを示す値として寄与率を用いる。正準変量をいくつまで取り出せばいいかを判断する目安になる。

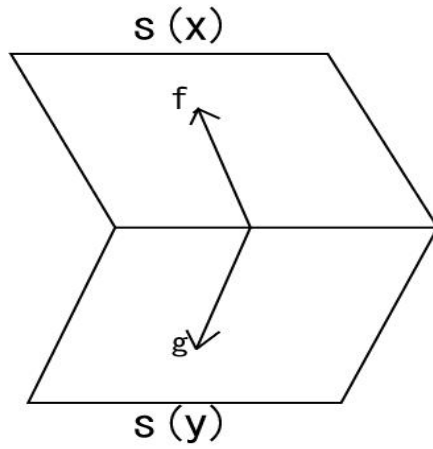


図 4.1: 正準相関分析のイメージ

4.2 正準相関分析の定式化

線形結合 f, g の相関係数

$$\rho(a, b) = \frac{\sigma_{fg}}{\sqrt{\sigma_f \sigma_g}} = \frac{\vec{a}' \Sigma_{YX} \vec{b}}{\sqrt{(\vec{a}' \Sigma_{YY} \vec{a})(\vec{b}' \Sigma_{XX} \vec{b})}}$$

を最大にするように \vec{a}, \vec{b} を求める。ここで $\Sigma_{ij}, (i, j = X, Y)$ は X, Y の分散共分散行列 Σ

$$\Sigma = \begin{pmatrix} \sigma_1 & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_2 & \cdots & \sigma_{2n} \\ \vdots & & & \vdots \\ \sigma_{n1} & \cdots & & \sigma_{nn} \end{pmatrix}$$

($n=q+p$) により

$$\Sigma = \begin{pmatrix} \Sigma_{XX} & \Sigma_{XY} \} q \\ \underbrace{\Sigma_{YX}}_q & \underbrace{\Sigma_{YY}}_p \} p \end{pmatrix}$$

である。

これを解くには

$$\vec{a}' \Sigma_{YY} \vec{a} = 1, \vec{b}' \Sigma_{XX} \vec{b} = 1$$

と基準化条件をかし、ラグランジュの未定乗数法を適用して

$$\phi_1 = \vec{a}' \Sigma_{YX} \vec{b} - \frac{1}{2} \lambda (\vec{a}' \Sigma_{YY} \vec{a} - 1) - \frac{1}{2} \theta (\vec{b}' \Sigma_{XX} \vec{b} - 1)$$

を最大にする \vec{a}, \vec{b} を求めればよい。 \vec{a} と \vec{b} について偏微分して 0 とおけば

$$-\lambda \Sigma_{YY} \vec{a} + \Sigma_{YX} \vec{b} = 0 \quad (4.1)$$

$$\Sigma_{XY} \vec{a} - \theta \Sigma_{XX} \vec{b} = 0 \quad (4.2)$$

上式に左からそれぞれ \vec{a}, \vec{b} をかけ、基準化条件を用いると

$$\lambda = \theta = \vec{a}' \Sigma_{YX} \vec{b} = \rho(a, b)$$

が成り立つ。

(2) 式の左から Σ_{XX}^{-1} を乗ずると

$$\lambda \vec{b} = \Sigma_{XX}^{-1} \Sigma_{XY} \vec{a}$$

これを (1) 式に代入して \vec{b} を消去することによって方程式

$$(\Sigma_{YX} \Sigma_{XX}^{-1} \Sigma_{XY} - \lambda^2 \Sigma_{YY}) \vec{a} = 0 \quad (4.3)$$

がえられ、同様にして \vec{b} に関する方程式

$$(\Sigma_{XY} \Sigma_{YY}^{-1} \Sigma_{YX} - \lambda^2 \Sigma_{XX}) \vec{b} = 0 \quad (4.4)$$

が得られ固有値問題に帰着する。(3) 式と (4) 式は共通の固有値

$$\lambda_1^2 \geq \lambda_2^2 \geq \dots \geq \lambda_r^2 \geq 0, r = \min(q, p)$$

を持つ。 λ は相関係数を表し、これを最大にすることが求められているので λ_1^2 をとりこれに対応する固有ベクトルを基準化して \vec{a}_1, \vec{b}_1 を求め、これらを用いて線形結合

$$f_1 = \vec{a}'_1 \vec{X}$$

$$g_1 = \vec{b}'_1 \vec{Y}$$

を構成する。 λ_1 を第 1 正準相関係数, (f_1, g_1) を第 1 正準変量と言う。

2 組の変数のグループ間の相関関係が、第 1 正準変量だけで十分説明できないときには、2 番目の正準変量を考える。第 1 正準変量と無相関つまり直交するようにし、相関が最大になるように新しい線形結合を作る。ラグランジュの未定乗数法をもちいて定式化すると (3) (4) の固有値問題が導かれる。 λ_1 は第 1 正準相関係数に用いているので λ に対応する固有ベクトル $\{\vec{a}_2, \vec{b}_2\}$ を用いて線形結合 f_2, g_2 を作る。これを第 2 正準変量といい、 λ_2 を第 2 正準相関係数という。

第 2 正準変量でも十分でないなら、同様にして第 3 正準変量, 第 4 正準変量, ... を求める。

こうして求めた正準変量 f_i と元の変数 \vec{X} との共分散 $\sigma_{f_i \vec{X}}$ を \vec{X} の f_i への正準負荷量といい、各変量はその正準変量によってどのくらい代表されているかを表している。

正準相関分析において，大きな正準相関が得られた場合にいつでも有用な情報が引き出せるとは限らない。 \vec{Y} と \vec{X} の 1 個や 2 個の成分間の特殊な関係だけを示していて，より一般的な構成間関係を必ずしも示していないかもしれないからである。個々の成分 x_j, y_j ごとに正準変数 f_i, g_i との共分散を負荷量として計算し評価する必要がある。 f_i, g_i がそれぞれ \vec{X}, \vec{Y} と全体としてもつ関連性の度合いを測るものとして寄与率

$$C_{f_i, \vec{X}}^2 = \frac{1}{q} \sum_{j=1}^q (\sigma_{f_i x_j})$$

$$C_{g_i, \vec{Y}}^2 = \frac{1}{p} \sum_{j=1}^p (\sigma_{g_i y_j})$$

が計算される。 $C_{f_i, \vec{X}}^2, C_{g_i, \vec{Y}}^2$ のことを，それぞれ f_i の \vec{X} への， g_i の \vec{Y} への寄与率という。

4.3 正準相関分析の例

正準相関分析の例として高校 1 年生の体力診断テスト，運動能力テストの結果である。まず 7 種類の体力診断テスト（第 1 変量群）とデータを示す。

生徒番号	反復横跳び (回)	垂直跳び (cm)	背筋力 (kg)	握力 (kg)	踏み台昇降 (指数)	立位体前屈 (cm)	伏臥上体そらし (cm)
1	46	55	126	51	75.0	25	72
2	52	55	95	42	81.2	18	50
3	46	69	107	38	98.0	18	74
4	49	50	105	48	97.6	16	60
5	42	55	90	46	66.5	2	68
6	48	61	106	43	78.0	25	58
7	49	60	100	49	90.6	15	60
8	48	63	122	52	56.1	17	68
9	45	55	105	48	76.0	15	61
10	48	64	120	38	60.2	20	62
11	49	52	100	42	53.4	6	42
12	47	62	100	34	61.2	10	62

生徒番号	反復横跳び (回)	垂直跳び (cm)	背筋力 (kg)	握力 (kg)	踏み台昇降 (指数)	立位体前屈 (cm)	伏臥上体そらし (cm)
13	41	51	101	53	62.4	5	60
14	52	55	125	43	86.3	5	62
15	45	52	94	50	51.4	20	65
16	49	57	110	47	72.3	19	45
17	53	65	112	47	90.4	15	75
18	47	57	95	47	72.3	9	64
19	48	60	120	47	86.4	12	62
20	49	55	113	41	84.1	15	60
21	48	69	128	42	47.9	20	63
22	42	57	122	46	54.2	15	63
23	54	64	155	51	71.4	19	61
24	53	63	120	42	56.6	8	53
25	42	71	138	44	65.2	17	55
26	46	66	120	45	62.2	22	68
27	45	56	91	29	66.2	18	51
28	50	60	120	42	56.6	8	57
29	42	51	126	50	50.0	13	57
30	48	50	115	41	52.9	6	39
31	42	52	140	48	56.3	15	60
32	48	67	105	39	69.2	23	60
33	49	74	151	49	54.2	20	58
34	47	55	113	40	71.4	19	64
35	49	74	120	53	54.5	22	59
36	44	52	110	37	54.9	14	57
37	52	66	130	47	45.9	14	45
38	48	68	100	45	53.6	23	70

次に5種類の運動能力テスト(第2変量群)の結果を示す。

生徒番号	50 m 走(秒)	走り幅跳び(cm)	遠投(m)	懸垂(回)	持久走(秒)
1	6.8	489	27	8	360
2	7.2	464	30	5	348
3	6.8	430	32	9	386
4	6.8	362	26	6	331
5	7.2	453	23	11	391
6	7.0	405	29	7	389
7	7.0	420	21	10	379
8	7.1	466	28	2	362
9	6.8	415	24	6	386
10	7.1	413	28	7	398
11	7.4	404	23	6	400
12	7.2	427	25	7	407
13	8.0	372	25	3	409
14	6.8	496	30	10	350
15	7.6	394	24	3	399
16	7.0	446	30	11	337
17	6.6	446	30	12	357
18	6.6	420	25	4	447
19	6.8	447	28	11	381
20	7.0	398	27	4	387
21	7.1	485	30	7	350
22	7.2	400	28	6	388
23	6.9	511	33	12	298
24	7.5	430	29	4	353
25	7.0	487	29	9	370
26	7.4	470	28	7	360
27	7.9	380	26	5	358
28	6.8	460	32	5	348
29	7.7	398	27	2	383
30	7.4	415	28	6	314
31	6.9	470	27	11	348
32	7.6	450	28	10	326
33	7.0	500	30	12	330

生徒番号	50 m 走 (秒)	走り幅跳び (cm)	遠投 (m)	懸垂 (回)	持久走 (秒)
34	7.6	410	29	7	331
35	6.9	500	33	21	348
36	7.5	400	29	2	421
37	6.8	505	28	11	355
38	7.2	522	28	9	352

以上のデータから相関行列を決め、正準変量を求める。求めた正準変量を以下の表で示す。

表 4.1: 正準変量の係数

	第 1 正準変量 f_1	第 2 正準変量 f_2
反復横跳び	0.44214	-0.20878
垂直跳び	0.26687	0.70210
背筋力	0.58837	-0.21022
握力	0.06141	0.01477
踏み台昇降	0.22167	-0.72625
立位体前屈	0.09112	-0.17485
伏臥上体そらし	0.01385	0.23987
	第 1 正準変量 g_1	第 2 正準変量 g_2
50 m 走	-0.42660	0.82554
走り幅跳び	0.23347	1.04037
ボール投げ	0.36966	0.19830
懸垂	0.00372	0.22192
持久走	-0.35605	0.81011
正準相関係数	0.85147	0.72841

表 4.2: 正準負荷量

	第 1 正準変量 f_1	第 2 正準変量 f_2
反復横跳び	0.67612	-0.26416
垂直跳び	0.65072	0.61202
背筋力	0.74747	0.17035
握力	0.23826	0.02665
踏み台昇降	0.16764	-0.72979
立位体前屈	0.37694	0.12045
伏臥上体そらし	0.10655	0.17351
	第 1 正準変量 g_1	第 2 正準変量 g_2
50 m走	-0.65611	0.26984
走り幅跳び	0.77604	0.52846
ボール投げ	0.77604	0.52846
懸垂	0.62403	0.20140
持久走	-0.69130	0.18738

踏み台昇降の指数は大きいほど体力の高さを表し、第 2 変量群 (運動能力テスト) の 1 番目 (50 m走) と 5 番目 (持久走) は、一定距離を走るのに要する時間で測定されているので値が小さいほど運動能力が高いことを表し、他の変量と異なることに注意しこれらの係数、負荷量の符号を逆転して解釈を試みる。

第 1 正準変量はそれぞれの変量群内で、どのテストの結果がよくても値が大きくなるので大まかには全般的な能力の高さを表わすといえるが、第 1 変量群 (体力診断テスト) では、とくに反復横とび、背筋力の係数 (負荷量) が大きく、第 2 変量群 (運動能力テスト) では、とくに 50m 走、ボール投げ、持久走の係数 (負荷量) が大きいので、それぞれこれらの能力の高さを代表する正準変量と解釈される。

第 2 正準変量では、第 1 変量群は正準負荷量より垂直とびと踏み台昇降の寄与が大きくそれぞれ正と負になっており、第 2 変量群では走り幅跳びとボール投げが大きく 50 m走と持久走の走力の項目は小さくなっているため、おおまかには全般的な能力が同じ中で、跳ぶ能力が高いか走る能力が高いかを表す正準変量と解釈される。

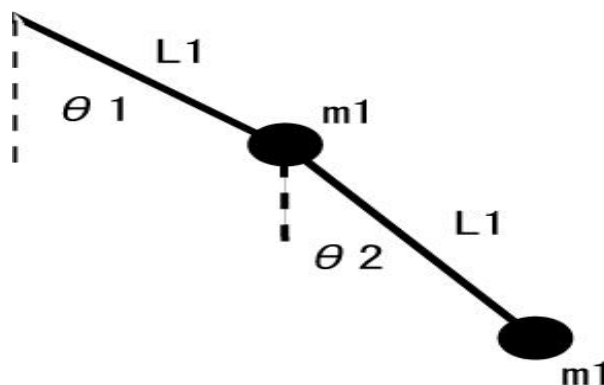
第5章 数値計算

5.1 主成分分析の結果

今回の研究では力学系として二重振り子を用いる。2重振り子はラグランジアン

$$L(\theta_1, \theta_2, \dot{\theta}_1, \dot{\theta}_2) = \frac{m_1 + m_2}{2} L_1^2 \dot{\theta}_1^2 + \frac{m_2}{2} L_2^2 \dot{\theta}_2^2 + m_2 L_1 L_2 \dot{\theta}_1 \dot{\theta}_2 \cos(\theta_1 - \theta_2) \\ + (m_1 + m_2) g L_1 \cos \theta_1 + m_2 g L_2 \cos \theta_2$$

を持つ。ここで g は重力加速度, L_1, L_2 は棒の長さ m_1, m_2 は質点の質量である。エネルギーが一定なので3つの自由度をもちカオス的なふるまいをする。



今回の計算ではパラメータとして, $L_1 = L_2 = 1[m], m_1 = m_2 = 1[kg], g = 9.8[m/s^2]$, 初期値は $\theta_1 = -\frac{17}{18}\pi[rad], \theta_2 = \frac{5}{6}\pi[rad], \dot{\theta}_1 = 2[rad/s], \dot{\theta}_2 = 0[rad/s]$ を用いた。

2重振り子の変数を4次のルンゲクッタ法で時間の刻み dt を 0.001 として数値計算し1つの時系列データとして $\dot{\theta}_1(t)$ を用いて, この変数から等しい時間間隔 Δt を持つ10次元の時間遅れベクトル

$$v(t) = (\dot{\theta}_1(t), \dot{\theta}_1(t + \Delta t), \dots, \dot{\theta}_1(t + 9\Delta t))$$

を作る。

$t = 0$ から $t = (m - 1)\Delta t$ まで t が Δt ずつ増えるように m 個の時間遅れベクトルを作る。

$$v_1(t) = (\dot{\theta}_1(0), \dot{\theta}_1(\Delta t), \dots, \dot{\theta}_1(9\Delta t))$$

$$v_2(t) = (\dot{\theta}_1(\Delta t), \dot{\theta}_1(2\Delta t), \dots, \dot{\theta}_1(10\Delta t))$$

⋮

$$v_m(t) = (\dot{\theta}_1((m-1)\Delta t), \dot{\theta}_1(m\Delta t), \dots, \dot{\theta}_1((m+8)\Delta t))$$

それぞれの時間遅れベクトルの第 1 成分, 第 2 成分, ..., 第 10 成分を 1 つのグループとする。つまり m 個の要素からなる 10 組のグループ

$$(\dot{\theta}_1(0), \dot{\theta}_1(\Delta t), \dots, \dot{\theta}_1((m-1)\Delta t)),$$

$$(\dot{\theta}_1(\Delta t), \dot{\theta}_1(2\Delta t), \dots, \dot{\theta}_1(m\Delta t)),$$

⋮

$$(\dot{\theta}_1(9\Delta t), \dot{\theta}_1(10\Delta t), \dots, \dot{\theta}_1((m+8)\Delta t))$$

を作り, それらの各グループを 10 個の変量として主成分分析する。今回は $m=4000$ で計算する。時間間隔 Δt を変えこれを行い時間間隔の変化に対して主成分分析した時の情報の縮約の変わり方を調べ, 2 重振り子は 3 次元なので第 3 主成分までの累積寄与率がより大きくなるような Δt を求める。

$\Delta t=1$ のときの主成分分析の結果

$\Delta t=1$ のときは、特に大きな寄与率はない。これから各主成分に情報が分散し 10 次元空間に広がっていることがわかる。これはアトラクターがうまく再構成できていないことを示している。

表 5.1: $\Delta t=1$ のときの主成分分析の寄与率

	寄与率	累積寄与率
第 1 主成分	0.1666	0.1666
第 2 主成分	0.1526	0.3212
第 3 主成分	0.1232	0.4445
第 4 主成分	0.1035	0.5480
第 5 主成分	0.0920	0.6400
第 6 主成分	0.0899	0.7299
第 7 主成分	0.0803	0.8102
第 8 主成分	0.0695	0.8797
第 9 主成分	0.0613	0.9410
第 10 主成分	0.0590	1.000

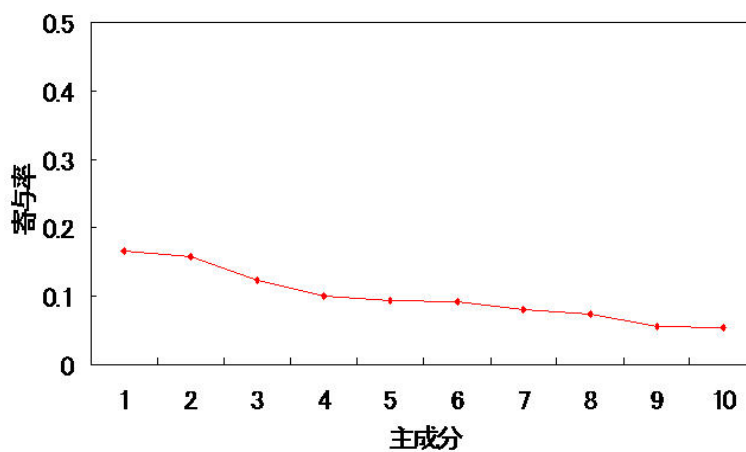


図 5.1: $\Delta t=1$ のときの主成分分析の結果

$\Delta t=0.02$ のときの主成分分析の結果

$\Delta t=0.02$ の時は, 時間間隔が小さすぎアトラクターが潰れてしまい, 情報が第 1 主成分に集中してしまいうまく再構成できない。

表 5.2: $\Delta t=0.02$ のときの主成分分析の寄与率

	寄与率	累積寄与率
第 1 主成分	0.8942	0.8942
第 2 主成分	0.0664	0.9605
第 3 主成分	0.0162	0.9767
第 4 主成分	0.0077	0.9844
第 5 主成分	0.0046	0.9890
第 6 主成分	0.0032	0.9922
第 7 主成分	0.0024	0.9946
第 8 主成分	0.0020	0.9966
第 9 主成分	0.0018	0.9984
第 10 主成分	0.0016	1.000

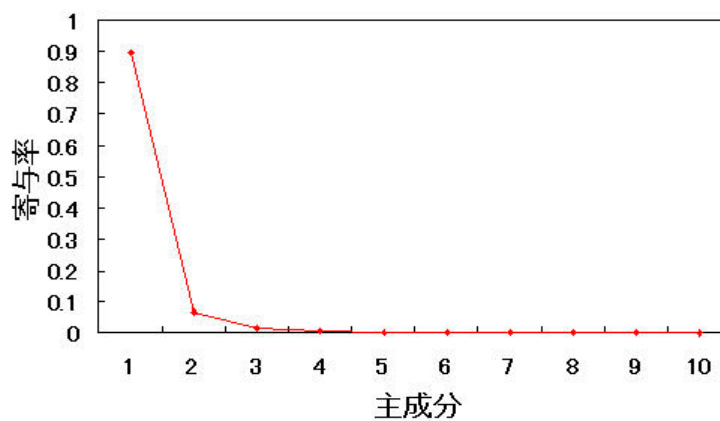


図 5.2: $\Delta t=0.02$ のときの主成分分析の結果

$\Delta t=0.1$ のときの主成分分析の結果

$\Delta t=0.1$ のときは第1・第2主成分の寄与率が特に大きく、第3主成分までの累積寄与率が約0.87であり、3次元まででほとんどの情報が表せている。二重振り子の自由度は3なので、アトラクターの再構成に成功したといえる。

表 5.3: $\Delta t=0.1$ のときの主成分分析の寄与率

	寄与率	累積寄与率
第1主成分	0.4843	0.4843
第2主成分	0.2965	0.7808
第3主成分	0.0902	0.8710
第4主成分	0.0446	0.9156
第5主成分	0.0251	0.9407
第6主成分	0.0168	0.9576
第7主成分	0.0128	0.9704
第8主成分	0.0108	0.9812
第9主成分	0.0096	0.9908
第10主成分	0.0092	1.000

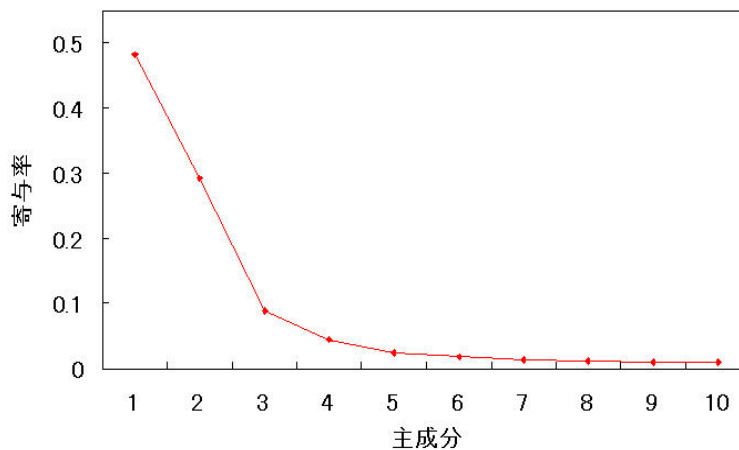


図 5.3: $\Delta t=0.1$ のときの主成分分析の結果

主成分の係数

$\Delta t=0.1$ のときの、各主成分の係数について調べる。第 1, 第 2, 第 3 主成分の各係数の値は以下のようにになっている。

	第 1 主成分	第 2 主成分	第 3 主成分
第 1 成分	0.41	0.70	0.43
第 2 成分	0.57	-0.69	0.33
第 3 成分	0.71	-0.58	0.09
第 4 成分	0.81	-0.39	-0.17
第 5 成分	0.87	-0.14	-0.34
第 6 成分	0.87	0.14	-0.34
第 7 成分	0.81	0.39	-0.17
第 8 成分	0.71	0.59	0.09
第 9 成分	0.57	0.69	0.33
第 10 成分	0.40	0.70	0.43

これをグラフに表す。

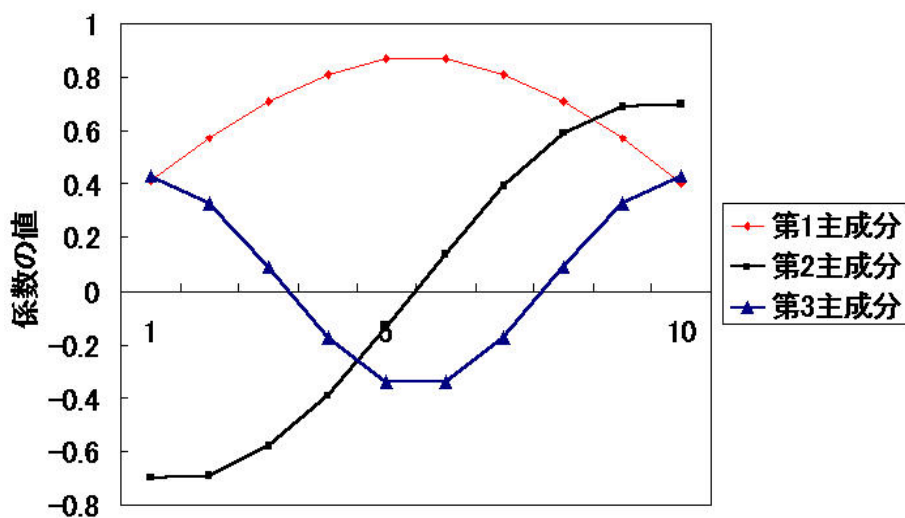


図 5.4: 主成分の係数

第 1 主成分のグラフをみると、各係数の値は差は大きくなく平均 0.62 であり、緩やかな曲線である。このことと時間遅れベクトルが θ_1 から構成されていることを考慮すると、第 1 主成分は θ_1 の平均に比例しているのではないかと推測できる。

またこのグラフから、第 1, 第 3 主成分の係数は左右に対称になっており中心にそれぞれ極大、極小を持ち、第 2 主成分は単調増加し、中心が 0 であることがわかる。このことから第 2, 第 3 主成分は第 1 主成分のそれぞれ 1 階, 2 階微分になっている。つまり第 2, 第 3 主成分は θ_1 の 2 階微分, 3 階微分になっている。以下まとめると

$$\text{第 1 主成分} \simeq \frac{\dot{\theta}_1(t) + \dot{\theta}_1(t+1)}{2}$$

$$\text{第 2 主成分} \simeq \frac{-\dot{\theta}_1(t) + \dot{\theta}_1(t-1)}{2} = \ddot{\theta}_1$$

$$\text{第 3 主成分} \simeq \frac{\dot{\theta}_1 + 2\dot{\theta}_1(t-1) + \dot{\theta}_1(t-2)}{4} = \frac{d\ddot{\theta}_1}{dt}$$

5.2 正準相関分析の結果

$\Delta t = 0.1$ の時間遅れベクトルを主成分分析して得られた第 1 主成分, 第 2 主成分, 第 3 主成分を 1 つのグループとし, また計算から得られた 2 重振り子の変数 $\theta_1, \theta_2, \dot{\theta}_1, \dot{\theta}_2$ をもう 1 つのグループとして正準相関分析する。正準相関係数を計算し正準変量のうち正準相関係数の大きいものの正準負荷量と寄与率を調べ 3 つの主成分が元の変数からどのように構成されているかを調べる。

表 5.4: 正準相関係数

	第 1 正準変量	第 2 正準変量	第 3 正準変量
正準相関係数	0.921872	0.309903	0.164999

第 3 正準変量の相関係数は小さいので第 1, 第 2 正準変量の正準負荷量を調べる。

表 5.5: 正準変量の係数

元の変数	f_1 (第 1 正準変量)	f_2 (第 2 正準変量)
θ_1	-0.55165	-0.032142
θ_2	0.018818	0.032422
$\dot{\theta}_1$	-0.019372	0.287778
$\dot{\theta}_2$	0.003534	0.070793
主成分値	g_1	g_2
第 1 主成分	-0.210100	0.402087
第 2 主成分	0.436266	0.246804
第 3 主成分	-0.496027	-0.200921

表 5.6: 正準負荷量

元の変数	f_1 (第 1 正準変量)	f_2 (第 2 正準変量)
θ_1	-1.800339	-0.100897
θ_2	0.117689	0.104172
$\dot{\theta}_1$	-0.229763	3.177674
$\dot{\theta}_2$	0.062693	1.114706
主成分値	g_1	g_2
第 1 主成分	-1.017517	1.947310
第 2 主成分	1.293535	0.731776
第 3 主成分	-0.447344	-0.181202

係数と正準負荷量より第 1 正準変数 f_1 は θ_1 の寄与が大きい。これに対応する g_1 は

$$g_1 \simeq -2(\text{第 1 主成分}) + 4(\text{第 2 主成分}) - 4(\text{第 3 主成分}) \simeq -2\dot{\theta}_1 + 4\ddot{\theta} - 4\frac{d\ddot{\theta}_1}{dt}$$

となっている。時間遅れベクトルは $\dot{\theta}_1$ から構成されているが f_1 は θ_1 だけでほとんど表すことができる。また第 2 正準変量 f_2, g_2 は

$$f_2 \simeq 3\dot{\theta}_1 + \dot{\theta}_2$$

$$g_2 \simeq 2(\text{第 1 主成分}) + (\text{第 2 主成分}) \simeq 2\dot{\theta}_1 + \ddot{\theta}_1$$

と表せる。

第6章 まとめ

今回の研究では、主成分分析を用いて2重振り子のアトラクターを1変数から再構成し次元を推定することに成功したといえるのではないか。また正準相関分析の結果からは時間遅れベクトルに用いた $\dot{\theta}_1$ が必ずしも第1正準変量により強く反映されているわけではない、という興味深い結果になった。

しかし当然ながら今回の結果から全ての力学系でこの方法が有効だとはいえない。この手法がより実用的な手法というためには、より多くの実証や数学的な裏づけが必要である。

関連図書

- [1] 相原 一幸 (編)「カオス時系列解析の基礎と応用」(2000)
- [2] 塩谷 實 「多変量解析概論」(1990)
- [3] Floris Takens. "Detecting strange attractor in turbulence". In D.A.Rand and B.S Young, editors, "Dynamical Systems of Turbulence", Vol.898 of Lecture Notes in Mathematics, pp.366-381, Berlin, 1981. Springer-Verlag.